

# Histograma de gradientes para poses de mão em um ambiente automotivo

Gustavo Müller Nunes

June 2014

# Sumário

<b>1</b>	<b>Introdução</b>	<b>4</b>
1.1	Apresentação . . . . .	4
1.2	Objetivo . . . . .	5
1.3	Justificativa . . . . .	5
1.4	Hipótese . . . . .	6
1.5	Metodologia . . . . .	6
1.6	Organização da dissertação . . . . .	7
<b>2</b>	<b>Referencial Teórico</b>	<b>8</b>
2.1	Histograma orientado a gradientes . . . . .	8
2.1.1	Normalização Gamma/Cor . . . . .	8
2.1.2	Gradientes . . . . .	8
2.1.3	Cálculo dos histogramas . . . . .	11
2.1.4	Normalização em blocos . . . . .	11
2.2	Estado da arte . . . . .	11
<b>3</b>	<b>Pesquisa</b>	<b>14</b>
3.1	Especificação da base de pesquisa . . . . .	14
3.1.1	Construção da câmera infra vermelha . . . . .	14
3.1.2	Elaboração da base de dados . . . . .	15
3.2	Desenvolvimento da Pesquisa . . . . .	16
3.2.1	Implementação do HOG . . . . .	16
3.2.2	Resultados . . . . .	16
<b>4</b>	<b>Discussão</b>	<b>18</b>
<b>5</b>	<b>Conclusão</b>	<b>19</b>

# Lista de Tabelas

2.1	Parâmetros do HOG otimizado por Dalal . . . . .	12
3.1	Lista de usuários . . . . .	15
3.2	Parâmetros da base dados de referência . . . . .	16
3.3	Parametrização do conjunto 2 . . . . .	16
3.4	Parametrização do conjunto 3 . . . . .	17
3.5	Parametrização do conjunto 4 . . . . .	17

# Lista de Figuras

1.1	Kinect, da Microsoft, e a câmera da <i>Creative</i> com parceria da Intel . . . . .	4
1.2	Exemplo do uso de poses em um ambiente automotivo . . . . .	5
2.1	Fluxo de cálculo para extrair o vetor de características . . . . .	8
2.2	Exemplo de máscara 3x3 . . . . .	9
2.4	Gradientes . . . . .	10
3.1	Webcam utilizada na aquisição das imagens sem nenhuma modificação . . . . .	14
3.2	Lentes com o filtro infra vermelho localizado na parte traseira . . . . .	15

# Capítulo 1

## Introdução

### 1.1 Apresentação

Reconhecimento de gestos baseado em visão computacional é um assunto bastante pesquisado e já pode ser considerado popular, isto porque, a busca por mecanismos que tornem a interação entre homem e máquina mais intuitiva e natural é constante e vem aumentando com o lançamento de plataformas que auxiliam os desenvolvedores nos complexos algoritmos que envolvem essa área. O lançamento do Kinect, da Microsoft [18], e da plataforma de desenvolvimento da Intel, chamada Intel Perceptual Computing [19] (figura 1.1), ambas com câmeras de profundidade, vem popularizando o desenvolvimento de aplicativos e revolucionando o jeito que interagimos com os jogos e computadores.



Figura 1.1: Kinect, da Microsoft, e a câmera da *Creative* com parceria da Intel

O uso de câmeras em carros e caminhões também tem aumentando nos últimos anos. Sistemas de segurança capazes de verificar se o motorista esta saindo indevidamente da faixa, ou se o veículo esta em rota de colisão com algum outro automóvel ou objeto e até mesmo monitorando o stress do motorista já são comuns em vários modelos de veículos. Mas pouco vimos o uso dessas câmeras para interação do motorista com a grande quantidade de controles que temos nos carros. Sistemas de navegação, componentes de som e imagem como CD/DVD player, radio, televisão, celulares, computador de bordo, e ar condicionado são alguns exemplos de dispositivos que requerer uma constante interação com o motorista e cujo comandos poderiam ser dados por meio de gestos. O sistema de gestos também pode ser usado como um complemento ao sistema de reconhecimento de voz, bastante comum hoje nos carros.

Nesse sistema gestual de interação do motorista com o veículo, podemos entender poses e gestos como sendo movimentos ou poses executados pela mão direta do motorista dentro do campo de visão de uma câmera instalada no teto do carro. O nosso estudo, portanto, é focado no histograma orientado a gradientes (HOG - Histogram of oriented gradients) aplicado em um sistema de interface de usuário através de gestos. Um sistema em tempo real capaz de reconhecer poses de mão e gestos que permita o motorista interagir com o veículo de forma intuitiva e eficaz. Na figura 1.2 temos um exemplo de como seria uma pose de mão aberta dentro de um ambiente automotivo.

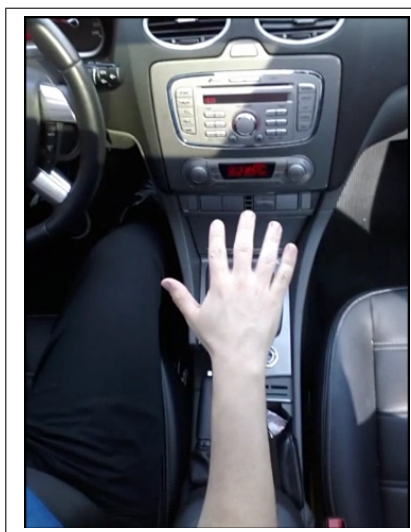


Figura 1.2: Exemplo do uso de poses em um ambiente automotivo

## 1.2 Objetivo

Um dos primeiros passos no processamento de uma imagem, com o objetivo de se identificar uma pose, seria encontrar a região de interesse. Essa região é uma sub-imagem onde a mão aparece com mais evidência, eliminando as partes que não são de interesse. A partir daí os mais diversos algoritmos podem ser aplicados para se identificar a pose. O objetivo do trabalho é encontrar essa região de interesse utilizando o HOG. HOG é um conjunto de técnicas que usam a orientação da variação das bordas para caracterizar a imagem. Uma grande vantagem dessa técnica é sua grande robustez contra variações de luminosidade. A proposta é portanto estudar o efeito da variação dos principais parâmetros do cálculo do HOG e assim encontrar qual a melhor configuração para a aplicação proposta. Um balanço entre performance e processamento deve ser levado em consideração, já que o trabalho computacional deve ser reduzido ao máximo para aplicações automotivas.

O HOG foi proposto por Dalal em 2005[17] que encontrou o melhor conjunto de parâmetros para a representação de seres humanos em diversas situações e poses diferentes. O que queremos avaliar é se para a nossa aplicação, os mesmos parâmetros podem ser aplicados e qual seria o custo na performance do algoritmo se o mesmo fosse simplificado com o intuito de reduzir processamento. O HOG poderia ser usado de duas maneiras diferentes: como um pre classificador para encontrar as regiões mais prováveis de se ter uma mão e assim limitar a imagem em algumas regiões de interesse aonde um segundo algoritmo seria aplicado, nesse caso não seria trabalho do HOG dizer qual é a pose, mas sim se é uma mão ou não, ou no máximo classificar a pose em algum grupo de poses (como feito em [10]). Mas o HOG poderia ser usado também para dizer qual pose é, sem a necessidade de nenhum algoritmo secundário. Visando que temos duas aplicações para o HOG, é possível que teremos duas configurações diferentes e portanto essas variações devem entrar no escopo desse trabalho.

## 1.3 Justificativa

A função principal do motorista deve ser sempre o controle do veículo. Distrações, como operar o rádio ou a central multimídia são exemplos constantes de cause de acidentes. Portanto apenas alguns poucos e curtos momentos podem ser usados para interagir com os comandos do veículo. Em estudos de usabilidade, o controle gestual provou ser mais intuitivo, efetivo [12] [13] e distrair menos do que o uso habitual de botões [14]. Por esse motivo, um estudo sobre técnicas para atingir esse objetivo é justificável.

As condições gerais dentro do automóvel inclui uma grande variação de iluminação, mudança de usuário (cor de pele, braço com ou sem vestimentas e vestimentas de cores e estampas diferentes) e fundos não uniformes. Além disso, a aceitação do usuário é um item bastante importante, portanto coisas como uma iluminação artificial visível, restrição de vestimentas e calibração extensiva não pode ser tolerados. Tendo isso em mente, alguns critérios e requisitos para o sistema podem ser estabelecidos:

- robustez contra ambientes ruidosos
- iluminação invisível
- independente de usuário
- sem calibração ou treinamento pelo usuário
- pequeno e compreensível conjunto de gestos
- reação do sistema com o mínimo de latência

Em 2005, Navneet Dalal [17] fez um estudo sobre histogramas orientado a gradientes aplicado à detecção de humanos. Seu estudo, variando cada parâmetro do cálculo dos histogramas e encontrando um conjunto de parâmetros que melhor servia para reconhecimento de humanos, virou referência para todos os estudos posteriores de HOG. Em seu texto ele diz que o uso de histogramas orientados tem muitos precursores [3] [4], mas que apenas atingiu a maturidade quando combinado com histogramas locais e normalização proposto pela Lowes Scale Invariant Feature Transformation (SIFT) [15]. A conclusão que ele chegou foi que usando histogramas de gradientes locais normalizados, similar ao SIFT, em um grade com sobreposição tem ótimos resultados para detecção de humanos, reduzindo falsos positivos em mais de uma ordem de magnitude comparado com Haar wavelets.

## 1.4 Hipótese

Trabalhamos com a hipótese de que o HOG pode ser parametrizado para encontrar poses de mão. O conjunto original de parâmetros podem ser modificados para melhor de adequar à aplicação proposta.

## 1.5 Metodologia

Nossa primeira etapa é a captura das poses, a pesquisa literária mostra ([2] e [1]) que o uso de uma câmera infravermelha simples já é adequado para o problema, onde o ambiente é iluminado por infravermelho de curta distância (950nm). A câmera ainda possui um filtro de luz, permitindo apenas que a luz infravermelha seja capturada pela câmera. Apesar de existir câmeras mais sofisticadas, optamos por usar a câmera mais simples, em vez das câmeras de profundidade por ser mais compatível com os padrões de mercado automotivo. No momento que esse texto foi escrito, as câmeras de profundidade ainda possuem um preço proibitivo e a quantidade de processamento é bastante limitada em um ambiente embarcado.

As imagens de poses e os vídeos dos gestos serão obtidos em dois ambientes distintos. Primeiro em um ambiente controlado com fundo homogêneo de cor preta e em uma sala totalmente escura (essa base de dados será usada como referência para os algoritmos implementados). O outro será obtido no interior de um veículo, tanto de dia como de noite. A captura das imagens no interior do veículo seria feita variando tanto o motorista quanto o veículo. Também será usado outras bases de dados que não em veículos para comparar a performance do algoritmo nas mais diversas situações.

Próximo passo seria implementar o algoritmo para que depois possamos variar os parâmetros e analisar com o melhor conjunto. Podemos usar como referência a implementação feito pelo MATLAB, que usou os parâmetros de Dalal.

## **1.6 Organização da dissertação**

Em construção.



## Capítulo 2

# Referencial Teórico

O objetivo desse capítulo é referenciar as teorias que embasam o conteúdo desse trabalho bem como o estado da arte no uso do HOG.

### 2.1 Histograma orientado a gradientes

HOG (Histogram of oriented gradients) é um descritor computado a partir dos gradientes da imagem, podemos defini-lo com sendo uma informação estatística do gradiente e intensidade de uma área. Suas principais propriedades são a robustez para pequenas variações nos locais dos contornos, direções e variações significativas na iluminação e cor. Na figura 2.1 temos um resumo das principais etapas do cálculo feito para extrair o vetor de características.

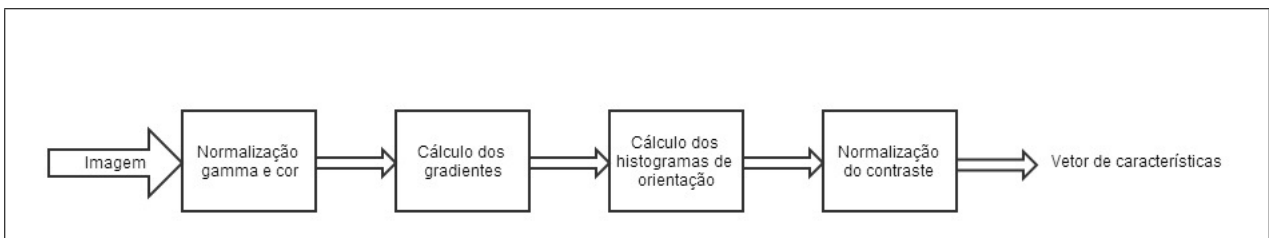


Figura 2.1: Fluxo de cálculo para extrair o vetor de características

Para usar como referência, daqui pra frente vamos nos referir ao conjunto de parâmetros do HOG usado pelo Dalal [17] para detecção de pessoas como sendo o HOG original.

#### 2.1.1 Normalização Gamma/Cor

Os pixels da imagem podem ser representados de diversas maneiras como escala de cinza, RGB e LAB. Uma normalização do gamma pode ainda ser aplicado. Apesar de imagens em tons de cinza apresentarem uma performance menor, essa será a opção de cor que iremos usar em nosso trabalho. O uso de câmeras infra vermelhas resulta na perda da informação de cor e portanto não teremos essa opção.

#### 2.1.2 Gradientes

Um dos mais importantes processos no processamento de uma imagem é a sua segmentação. A segmentação consiste em subdividir a imagem em regiões ou objetos de interesse. O nível de segmentação depende do problema a ser resolvido e é comumente baseado em duas propriedades do valor da intensidade: descontinuidade e similaridade. A primeira consiste em particionar uma imagem baseado nas mudanças abruptas na intensidade, como por exemplo as bordas de um objeto. Já na segunda, é feito

$w_1$	$w_2$	$w_3$
$w_4$	$w_5$	$w_6$
$w_7$	$w_8$	$w_9$

Figura 2.2: Exemplo de máscara 3x3

-1	-1	-1	-1	0	1
0	0	0	-1	0	1
1	1	1	-1	0	1

(a) Máscara Prewitt

-1	-2	-1	-1	0	1
0	0	0	-2	0	2
1	2	1	-1	0	1

(b) Máscara Sobel

o agrupamento de uma região baseado em sua similaridade com outras partes da imagem, como cor ou nível de intensidade.

Gonzales (ano do livro) define borda como sendo um conjunto de pixels conectados presente na fronteira entre duas regiões. E conclui que a magnitude da primeira derivada pode ser usada para detectar a borda em um ponto da imagem.

A derivada de primeira ordem de uma imagem digital pode ser aproximada no gradiente 2D. O gradiente de uma imagem  $f(x, y)$  no ponto  $(x, y)$  é definido como um vetor

$$\nabla \mathbf{f}(x, y) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2.1)$$

cuja magnitude é definida como  $\nabla f$ , onde

$$\nabla f = \text{mag}(\nabla \mathbf{f}) = [G_x^2 + G_y^2]^{1/2} \quad (2.2)$$

e a direção do vetor  $\alpha(x, y)$  sendo definida como

$$\alpha(x, y) = \tan^{-1} \left( \frac{G_y}{G_x} \right) \quad (2.3)$$

onde o ângulo é medido em referência ao eixo  $x$ . A direção de uma borda no ponto  $(x, y)$  é perpendicular à direção do vetor gradiente no ponto.

O cálculo dessas derivadas podem ser implementados usando máscaras como o da figura 2.2. A máscara é aplicada em cada pixel da imagem e um novo valor é calculado conforme a equação 2.4.

$$R = w_1 z_1 + w_2 z_2 + w_3 z_3 + \dots + w_9 z_9 = \sum_{i=1}^9 w_i z_i \quad (2.4)$$

Nas figuras 2.3a e 2.3b temos dois exemplo das máscaras mais utilizadas para cálculo de gradiente. Na figura 2.4 podem ver o resultado das máscaras em uma imagem de uma pose de mão aberta feita por uma câmera infra vermelha.

No HOG original a máscara usada é uma máscara centrada 1-D  $[-1 \ 0 \ 1]$ . O gradiente é computado da seguinte maneira.

$$G_x(x, y) = f(x + 1, y) - f(x - 1, y)$$

$$G_y(x, y) = f(x, y + 1) - f(x, y - 1)$$

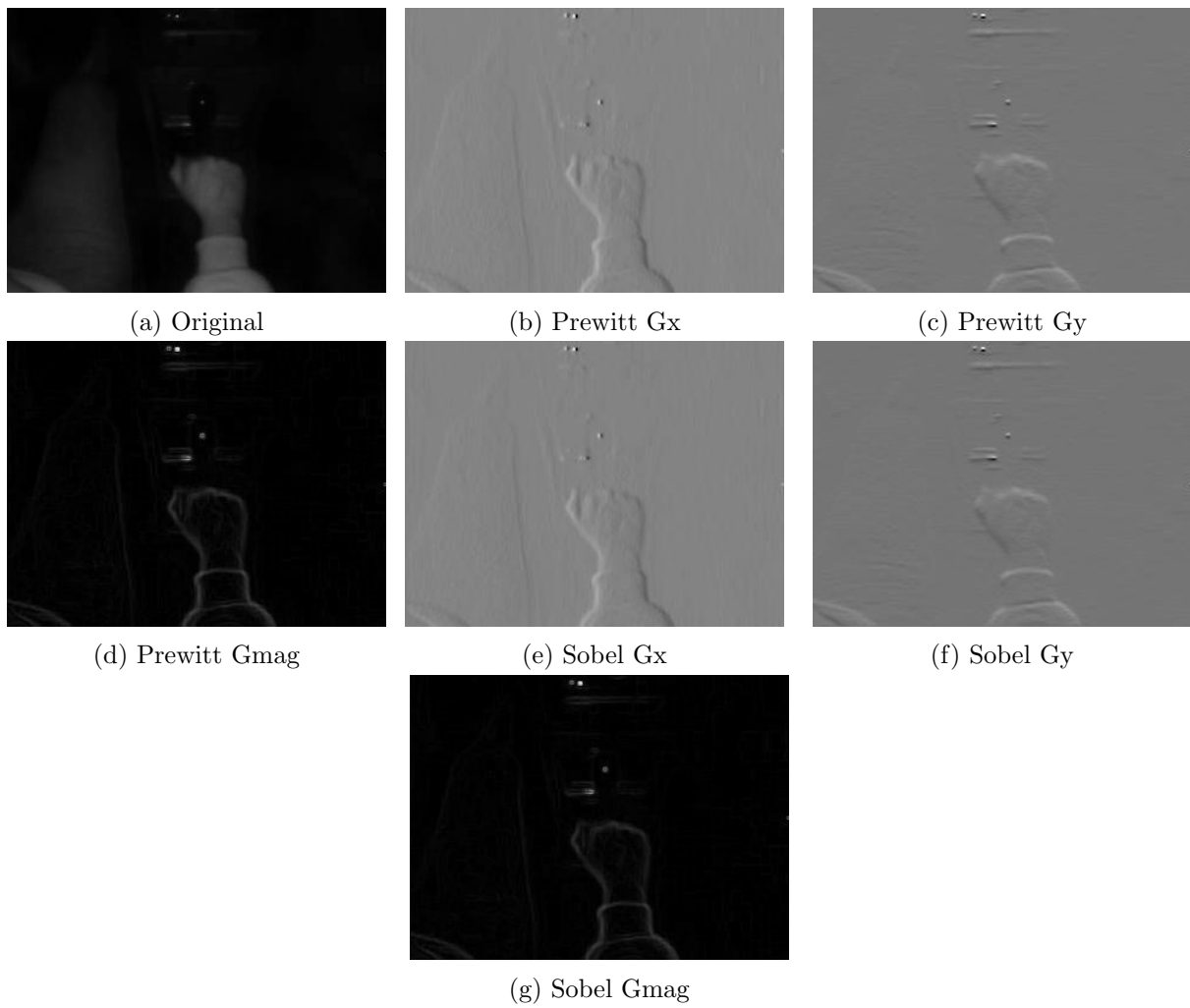


Figura 2.4: Gradientes

### 2.1.3 Cálculo dos histogramas

Depois dos cálculos do gradiente, a imagem é então dividida em pequenos retângulos (células). Para cada célula, um histograma é calculado. Esse histograma é a coleção dos ângulos dos vetores de gradiente de cada pixel que compõe a célula. Cada pixel apresenta um peso na construção do histograma das orientações das bordas. Esse peso pode ser em função do gradiente, do seu quadro ou da sua raiz.

Os ângulos podem ser agrupados variando de 0 à 360 graus ou de 0 à 180 graus.

No HOG original, as células tem tamanho 8x8, as orientações são ponderadas pela magnitude do vetor e uniformemente agrupas em 9 grupos de 0 a 180 graus.

### 2.1.4 Normalização em blocos

O tamanho do gradiente pode variar bastante por conta de variações como iluminação e contraste entre o fundo e o objeto de interesse. Portanto um importante passo para se obter um bom resultado na extração do vetor de características do objeto é a sua normalização.

Norma é uma função que atribui um tamanho de valor positivo e diferente de zero para um vetor em um espaço vetorial.

A função norma deve satisfazer algumas propriedades de escalabilidade e aditividade. Sendo um espaço vetorial  $V$  em um sub corpo  $F$  de números complexos, a norma em  $V$  é uma função  $p : \rightarrow \mathbf{R}$  com as seguintes propriedades.

- $p(a\mathbf{v}) = |a|p(\mathbf{v})$
- $p(\mathbf{u} + \mathbf{v}) \leq p(\mathbf{u}) + p(\mathbf{v})$
- Se  $p(\mathbf{v}) = 0$  então  $\mathbf{v}$  é o vetor zero.

Uma norma bastante usada é a norma euclidiana, que diz que em um espaço euclidiano  $R^n$  a norma será:

$$\|x\| := \sqrt{x_1^2 + \dots + x_n^2}$$

Dos vários esquemas de normalização, a maioria é baseada no agrupamento de células em blocos maiores e normalizando o contraste de cada bloco separadamente. Além disso a uma sobreposição entre blocos para que as células de cada bloco possa contribuir nas componentes de normalização diversas vezes. No HOG original extraiu-se o HOG em blocos de tamanho 16x16 e dividiu cada bloco em 4 células.

$$f(C_i, k) = \frac{\sum_{(x,y) \in C_i} V_k(x, y) + \varepsilon}{\sum_{(x,y) \in B} V_k(x, y) + \varepsilon}$$

$f(C_i, k)$  é a proporção do valor do gradiente acumulado do kth bin no bloco que contém a célula  $C_i$ . O  $\varepsilon$  é um valor bem pequeno para eliminar os denominadores zeros.

Depois cada histograma é concatenado, formando um vetor único de características.

O HOG proposto por Dalal [17] possui a seguinte parametrização conforme tabela 2.1.

## 2.2 Estado da arte

Um dos precursores em extração de características da mão usando histograma de orientação de gradientes (HOG) foi o laboratório da Mitsubishi que publicou um conjunto de artigos [3], [4] sobre o tema. Nesses artigos foi feito o HOG da imagem como um todo, em tons de cinza, e dividindo os ângulos em 36 grupos. O método não era geral o suficiente para ser usado com um algoritmo válido para representação de uma forma genérico e por isso, o mesmo evolui para o SIRF. Acontece que

Cor	RGB sem correção de gamma
Gradiente	$[-1, 0, 1]$ sem smoothing
Bins	9
Orientação	0 à 180
Tamanho do bloco	16x16 pixels
Tamanho da célula	8x8 pixels
Janela Gaussian	8 pixel
Normalização	L2-Hys
Janela de detecção	64x128

Tabela 2.1: Parâmetros do HOG otimizado por Dalal

a aplicação é bastante parecida com a proposta desse trabalho e por isso HOG é melhor detalhado posteriormente.

Em [2] (Alemanha, 2000) e em [1] (Alemanha, 2003) temos um cenário idêntico ao proposto, onde imagens infra vermelhas de uma câmera instalada no teto do carro são capturadas e traduzidas em gestos e poses de mão. Em [1] o sistema proposto pelo artigo é capaz de reconhecer onze gestos e quatro poses. A imagem capturada em uma taxa de 25 fps e resolução 384x144 é primeiramente processada com uma combinação de subtração de fundo e threshold global. Em [2] é usado apenas um threshold global. A mão é considerada o maior objeto da cena. Depois da segmentação, um filtro para retirar o braço é aplicado e finalmente são calculados os momentos da imagem, para o cálculo da área e do centro de massa, e os momentos Hu.

Em [5] (Estados Unidos, 2008) temos também o uso de câmera infra vermelha no teto do carro, mas com o objetivo de discriminar qual pessoa está usando o painel de controles do carro, o motorista ou o passageiro. O sistema faz uso do HOG para descrever a imagem e um classificador SVM com uma taxa de 96.8% de acerto. O cálculo do HOG é uma versão simplificada feita por [17]. Nesse artigo, depois de calculado o gradiente da imagem, a mesma é dividida em uma grade de células 2x2, o histograma é calculado para cada célula com 8 bins variando de 0 à 360 graus, portanto formando um vetor de 32 dimensões.

Em [6] (China, 2010) o HOG é utilizado para detectar ciclistas. No método proposto não é feito overlap no cálculo dos histogramas, como uma maneira de melhorar o tempo de processamento, e amostragem piramidal é utilizada para extrair características globais em diferentes escalas. As imagens utilizadas são em tons de cinza e um filtro gaussiano é aplicado antes do cálculo dos HOGs (contrariando as orientações do Dalal em [17]). O gradiente é calculado com máscara  $[-1 \ 0 \ +1]$ , os ângulos são calculados entre 0 e 180, e o histograma é dividido em 20 bins. A imagem é dividida em blocos de 16x16 sem divisão de células. O classificador utilizado é um SVM linear. Esse trabalho é interessante pois propõe um método para melhorar a velocidade do cálculo dos histogramas, o que pode ser útil para aplicações em tempo real embarcadas.

Um estudo comparando descritores locais, semi locais e globais é feito em [7] (França, 2011). O objetivo do trabalho é estudar qual seria o método mais adequado para descrever poses de mão em uma sala de cirurgia para que o médico possa enviar comandos para os aparelhos sem precisa encostar neles. Para descritores globais foi usado os momentos de Zernike (invariante em rotação, translação e escala) combinados com um classificador linear SVM. O HOG é usado como um descritor semi local e SIFT para locais. Apesar de não dar detalhes de como é feito os cálculos do HOG, o artigo mostra uma melhor performance do método.

Nesse artigo [16] (Espanha, 2011), o problema a ser resolvido era verificar, com o uso de uma câmera, se uma pessoa fez as seis diferentes poses de mão para o lavar correto das mãos. Primeiro as imagens são segmentadas por cor de pele e depois um estimador de posição do braço e da mão baseado em um filtro multi modal probabilístico é proposto. Um ROI é criado com o resultado do filtro e anterior e então HOG é aplicado, usando como classificador dois SVM independentes. Uma

para o HOG normal e outro para o HOF (Histogram of optical flow).

Nesse artigo [8] (Japão, 2012), é utilizado a coHOG (co-occurrence HOG) para reconhecimento de navios em imagens ISAR. No coHOG os blocos são agrupados em pares, aumentando a robustez para imagens em diferentes ângulos e na oclusão de algumas partes do navio. Por outro lado, o coHOG tem uma alta dimensão.

A abordagem desse artigo [10] (China, 2012) é selecionar, usando HOG e SVM, uma região de interesse para depois aplicar o filtro de cor de pele. O bloco tem tamanho 12x12 pixels com 2x2 células.

## Capítulo 3

# Pesquisa

### 3.1 Especificação da base de pesquisa

Nesse capítulo tem como objetivo descrever as etapas da pesquisa prática.

#### 3.1.1 Construção da câmera infra vermelha

A câmera utilizada nessa aplicação tem que ser capaz de capturar imagens nas mais diversas condições de luminosidade. Temos o caso, por exemplo, de um dia de sol cuja intensidade de luz é bem alta. Até o ponto onde não há luz nenhuma. Nesses casos é necessário uma iluminação própria, mas ao mesmo tempo, não pode atrapalhar o motorista. Por isso, a iluminação infra vermelha é muito utilizada. O custo é baixo e não interfere em nada no ambiente. O maior contratempo desse tipo de iluminação é que se perde toda a informação de cor. Para gerar a base de dados para o nosso estudo, utilizamos uma câmera normal de mercado, modificada para receber a luz infra vermelha e colocamos LEDs de infra vermelho para fazer a iluminação.

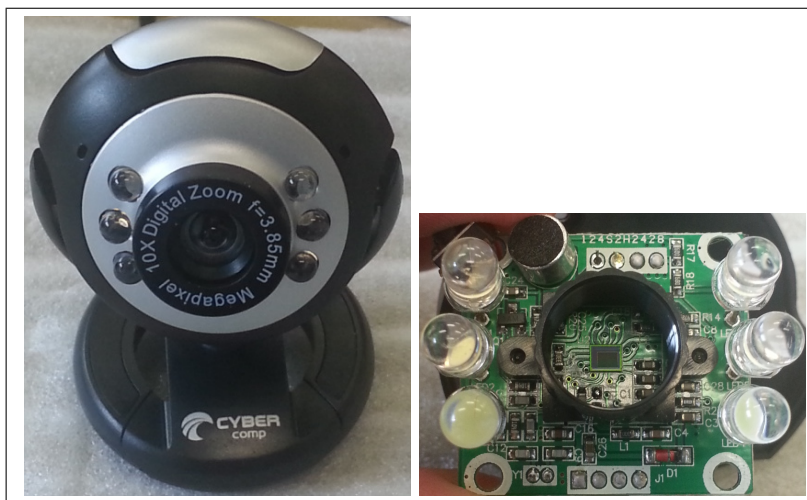


Figura 3.1: Webcam utilizada na aquisição das imagens sem nenhuma modificação

Na figura 3.1 temos a câmera utilizada para a aquisição das imagens. Nesse momento a câmera ainda não foi modificada. Essa câmera portanto ainda possui um filtro de luz infra vermelha e os LEDs de iluminação são LEDs brancos.

A principal modificação a ser feita nesse tipo de câmera é retirar o filtro infra vermelho. Esse filtro é uma placa de vidro localizado atrás da lente. Na figura 3.2 temos uma foto das lentes ainda com o filtro e depois já com o filtro retirado. E preciso também substituir os LEDs atuais, que são LEDs

brancos, para LEDs infra vermelho de 950nm.

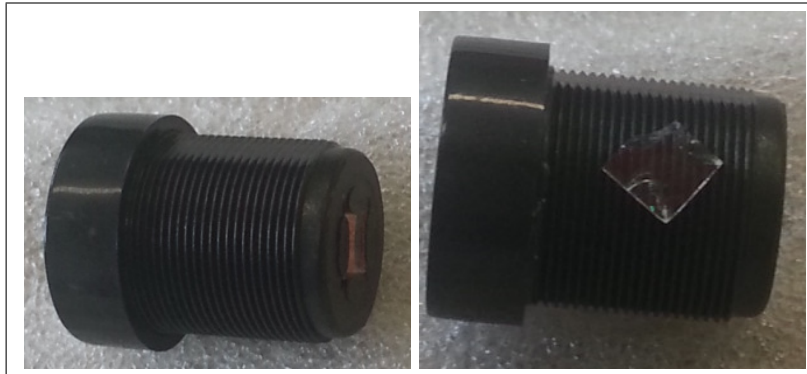


Figura 3.2: Lentes com o filtro infra vermelho localizado na parte traseira

### 3.1.2 Elaboração da base de dados

As bases de dados que serão usadas no trabalho precisam refletir as condições que encontramos em um ambiente automotivo. Por isso elaboramos um conjunto de banco de imagens que variam o fundo, o usuário, a iluminação e a vestimenta.

O nosso fundo vai variar conforme o carro aonde as imagens foram coletadas. Como referência, temos também um conjunto de imagens com o fundo preto homogêneo. O usuário será também modificado, variando sexo e cor de pele. A iluminação terá a captura diurna e noturna e a vestimenta varia por exemplo se o usuário esta usando blusa, relógio, etc.

Para o usuário temos a tabela 3.1 mostrando as principais características dos mesmos.

Usuário	Sexo	Cor de pele
1	Masculino	Branco
2	Masculino	Branco
3	Feminino	Branco
4	Masculino	Moreno
5	Feminino	Negra

Tabela 3.1: Lista de usuários

A nossa base de referência será uma banco de imagens com o fundo preto homogêneo, o usuário 1 do sexo masculino sem nenhum tipo de vestimenta ou acessório e a iluminação apenas dos LEDs infra vermelho, ou seja, em um ambiente totalmente escuro. Na tabela 3.2 temos um resumo da parametrização dessa base e alguns exemplos das imagens.

Na tabela 3.3 temos um conjunto de imagens com o fundo do carro Ford Focus, iluminação com LEDs infra vermelhos e usuário 1 com uma blusa preta.

Na tabela 3.4 temos um conjunto de imagens com o fundo do carro Ford Focus, iluminação com LEDs infra vermelhos e usuário 1 sem vestimentas.

Na tabela 3.5 temos um conjunto de imagens com o fundo do carro Passat, iluminação com LEDs infra vermelhos e usuário 2 usando uma blusa verde. O interessante desse conjunto é a existência de um LED no painel que pode atrapalhar a segmentação da imagem.



Usuário	Usuário 1
Fundo	Preto homogêneo
Iluminação	Infra vermelha
Vestimenta	Nenhuma




Gestos		
		

Tabela 3.2: Parâmetros da base dados de referência

Usuário	Usuário 1
Fundo	Ford Focus
Iluminação	Infra vermelha
Vestimenta	Blusa preta




Gestos		
		

Tabela 3.3: Parametrização do conjunto 2

## 3.2 Desenvolvimento da Pesquisa

### 3.2.1 Implementação do HOG

Em construção.

### 3.2.2 Resultados

Em construção.

Usuário	Usuário 1
Fundo	Ford Focus
Iluminação	Infra vermelha
Vestimenta	Nenhuma


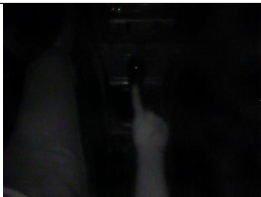
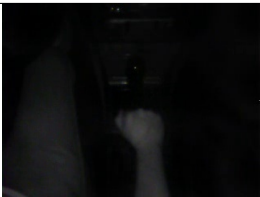
Gestos		
		

Tabela 3.4: Parametrização do conjunto 3

Usuário	Usuário 2
Fundo	Passat
Iluminação	Infra vermelha
Vestimenta	Blusa verde

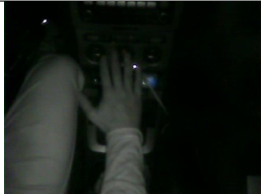
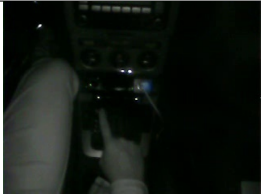
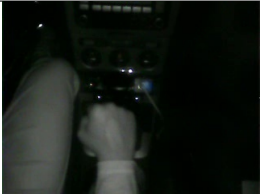
Gestos		
		

Tabela 3.5: Parametrização do conjunto 4

## Capítulo 4

# Discussão

O objetivo desse capítulo é discutir a relação entre a hipótese formulada no trabalho, a teoria existente sobre o assunto e a prática demonstrada no capítulo anterior.

## Capítulo 5

# Conclusão

Em construção.

# Referências Bibliográficas

- [1] Zobl, M., Nieschulz, R., Geiger, M., Lang M., Rigoll, G., Gesture Components for Natural Interaction with In-Car devices, 2003.
- [2] Akyol, S., Canzler, U., Bengler, K., Hahn, W.: Gesture control for use in auto-mobiles. In: Proceedings, MV A 2000 Workshop on Machine Vision Applications, Tokyo, Japan, November 28-30, 2000, IAPR, ISBN 4-901122-00-2 (2000)
- [3] W. T. Freeman and M. Roth. Orientation histograms for hand gesture recognition. Intl. Workshop on Automatic Face and Gesture-Recognition, IEEE Computer Society, Zurich, Switzerland, pages 296–301, June 1995.
- [4] W. T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma. Computer vision for computer games. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, pages 100–105, October 1996.
- [5] Shinko Y. Cheng and Mohan M. Trivedi. Real-time Vision-based Infotainment User Determination for Driver Assistance, IEEE Intelligent Vehicles Symposium, June 4-6, 2008
- [6] An Effective Crossing Cyclist Detection on a Moving Vehicle
- [7] Hand-gesture recognition: comparative study of global, semi-local and local approaches
- [8] Automatic Ship Recognition Robust Against Aspect Angle Changes and Occlusions
- [9] An Extended HOG Model: SCHOG for Human Hand Detection
- [10] A ROBUST METHOD OF FINGERTIP DETECTION IN COMPLEX BACKGROUND
- [11] Deformable HOG-based Shape Descriptor
- [12] Zobl, M., Geiger, M., Morguet, P., Nieschulz, R., Lang, M.: Gesture-based control of in-car devices. In: VDI-Berichte 1678: USEWARE 2002 Mensch-Maschine-Kommunikation/Design, GMA Fachtagung USEWARE 2002, Darmstadt, Germany, June 11-12, 2002, Dusseldorf, VDI, VDI-Verlag (2002) 305–309
- [13] Zobl, M., Geiger, M., Bengler, K., Lang, M.: A usability study on hand gesture controlled operation of in-car devices. In: Abridged Proceedings, HCI 2001 9th Int. Conference on Human Machine Interaction, New Orleans, Louisiana, USA, August 5-10, 2001, New Jersey, Lawrence Erlbaum Ass. (2001) 166–168
- [14] Geiger, M., Zobl, M., Bengler, K., Lang, M.: Intermodal differences in distraction effects while controlling automotive user interfaces. In: Proceedings Vol. 1: Usability Evaluation and Interface Design , HCI 2001 9th Int. Conference on Human Machine Interaction, New Orleans, Louisiana, USA, August 5-10, 2001, New Jersey, Lawrence Erlbaum Ass. (2001) 263–267

- [15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [16] "A vision-based system for automatic hand washing quality assessment"
- [17] Dalal, N. and Triggs, M, "Histograms of Oriented Gradients for Human Detection", in *Proc. Of IEEE CVPR2005*, vol. 1, pp. 886-893, June 2005.
- [18] <http://www.microsoft.com/en-us/kinectforwindows/develop/>
- [19] <http://software.intel.com/en-us/vcsourc/tools/perceptual-computing-sdk>